

Evaluation of Proposed and Existing Accelerated Research Programs by the Office of Naval Research

RONALD N. KOSTOFF

Abstract—A process to evaluate proposed and existing accelerated research programs is described. The key component of the process is the use of mixed level review panels, consisting of bench-level researchers, research managers, transitioning experts, technology experts, and Navy requirements experts, to review multiple programs (nominally) in technical areas of similar science. The panel members score specific research quality and mission relevance rating factors as well as an Overall Program Quality rating factor. As evidenced by multiple regression analyses, the research quality and mission relevance factors which the reviewers rated have been their dominant considerations in determining the Overall Program Quality rating factor score. This implies that rating factors which the reviewers felt were important were not omitted. The operational components of the review process are described in detail, and results of experiments in which program presentation, evaluation, and discussion times were varied are presented. Application experience of the process to Office of Naval Research (ONR) program evaluation is presented.

Index Terms—Research evaluation, research planning, peer review, expert panels, technology assessment.

I. INTRODUCTION

The objective of this paper is to describe a process which was developed to evaluate proposed and existing research programs. Sufficient detail of the process operation is provided to allow interested potential users to implement the entire process in their organizations. The rationale for instituting this process is discussed, and the methods which were used to validate selected process components are presented. The Office of Naval Research (ONR) experience with the applications of this process to existing and proposed programs is described.

Since the details of the ONR applications of the process require some understanding of the ONR structure and the ONR evaluation cycle, an overview of the ONR structure and evaluation cycle will be provided.

A. ONR Structure Overview

ONR is responsible for all basic and applied Naval research. ONR allocates its research funds to six internal Navy organizations, called claimants. After these claimants receive their research funds they, depending on their charter, may then contract with external organizations to perform research, they may perform research in-house, or they may do some combination of external funding and in-house research. It should be emphasized that the process discussed in this paper is relevant to the internal allocation of research funds among Navy claimants and not the subsequent competitive allocation of funds among performers. These claimants are the Contract Research Department (at ONR Headquarters), the Naval Research Lab, the Naval Ocean Research and Development Activity, the Applied Research Program (at ONR Headquarters), the Naval Facilities Engineering Command, and the Naval Medical Command. The bulk of the ONR budget for the claimants is divided between two types of programs. One type, called core, consists of long-term programs whose selection is at each claimant's discretion. Sixty percent of the

ONR budget goes toward core. The other type, called Accelerated Research Initiative (ARI), consists of promising research programs in which funding is concentrated for a limited duration (typically five years) in order to accelerate progress. All six claimants compete against each other in a structured competition to obtain funding for the ARI programs. All claimants and ONR Headquarters have their own internal methods for evaluating both proposed and existing core and ARI programs. Evaluation of proposed and some existing ARI programs by the Planning and Assessment Office (P&A) of ONR Headquarters will be described in detail in the following sections of the text.

II. PRESENT ONR EVALUATION PROCESS FOR PROPOSED AND EXISTING ARI PROGRAMS

While the ARI programs and their associated evaluation processes were initiated at ONR in 1981, the present ONR evaluation process contains components developed by the author at the U.S. Department of Energy (DOE) and were incorporated in the ongoing ONR process in 1983. After a short overview showing how the present evaluation process fits into the ONR planning cycle, the specific elements of the evaluation process will be presented.

A. Overview of Evaluation of Proposed ARI Programs

The ONR evaluation process for proposed programs runs from mid-fall to late spring, and culminates in recommended programs to be funded consistent with budget constraints. The evaluation process starts in mid-fall with the development of technical guidance (specification of promising areas of research which ONR recommends the claimants consider when formulating their research programs) to claimants. Each SEM¹ presents a written status report on his discipline. This report includes the identification of promising research opportunities for ONR. Using the SEM reports, the reports of panels (groups of experts in specific disciplines) run by the National Academy of Sciences to identify promising research opportunities for each technical discipline, and other research and Navy requirements documents, ONR develops a list of priority research areas (the technical guidance). This document is sent to the ONR claimants to provide technical guidance for their submittal of proposed programs for the competition. In late winter, claimants send brief technical descriptions of their proposed programs to ONR to initiate the process of selecting program evaluation panels (groups of experts in specific disciplines who will evaluate the proposed programs). The proposed programs are combined into areas of similar science, and panels, which will meet for one day per panel to evaluate these groups of programs, are identified by topic. The panels (and titles of programs in each panel) which were assembled in spring of 1985 are shown in Fig. 1.

Now the high *work load* intensity segment of the evaluation process begins. The next major step in the process is selection of reviewers for each panel.

Since the panels are asked to address research quality, transition potential, and Navy relevance when they evaluate each proposed program, panel members selected include experts in these three areas for each program. Typically, the research experts come from universities, government, and industry; transition experts come from the Office of Naval Technology and the Naval Systems Commands; and other relevance experts come from the Naval Systems Commands

Manuscript received November 2, 1987; revised March 1988. The review of this paper was arranged by Department Editor R. Balachandra.

The author is with the Office of Naval Research, Department of the Navy, Arlington, VA 22217-5000.

IEEE Log Number 8822518.

¹ All of ONR research is divided into 15 technical disciplines (physics, chemistry, mechanics, etc.), called subelements, for management purposes. Each subelement is assigned a technical expert (subelement monitor or SEM) from one of the ONR claimants whose function is to monitor the status of the discipline.

<u>MECHANICS/MATERIALS</u>		<u>LIFE SCIENCES</u>	<u>GENERAL PHYSICS</u>
SUPERSONIC COMBUSTION		MARINE BIOSURFACES	PLASMA MICROWAVE
MICROMECHANICS		MEMBRANE ELECTROCHEMISTRY	TERRESTRIALLY IMPORTANT SOLAR PHENOMENA
HIGH-STRENGTH FERROUS ALLOYS		SEPTIC SHOCK	PUMPING ATMOSPHERIC SPECIES
OCEAN FLOOR EMBEDMENT MECHANICS		HIGH PRESSURE NERVOUS SYNDROME MECHANISM	FREE ELECTRON LASER
SOLID-FLUID INTERFACES		CELLULAR EFFECTS OF EMF	
CAVITATION		VISUAL PERCEPTION	
<u>MATH/COMPUTER SCI</u>		<u>CHEMISTRY</u>	<u>OCEAN SCIENCES</u>
PROJECT OPTIMIZATION		CHEMISTRY ELECTRICAL OPTICAL MATERIALS	MARINE MICROLAYER
MILITARY SUPERCOMPUTING		SHOCK CHEMISTRY AND PHYSICS OF ENERGETIC MATERIALS	EXTENDED GEOSAT MISSION
ELECTRONIC WARFARE		IMPROVED CATALYTIC SYSTEMS	TROPICAL CYCLONE MOTION
SYMBOLIC COMPUTING		BORON PROPELLANT	POLAR SEAS
		GRADED INTERFACE CHEMISTRY	ARCTIC ACOUSTICS
			SURFACE ACOUSTIC INTERACTIONS
		<u>SOLID STATE PHYSICS</u>	
		MICROELECTRONIC STRUCTURES	
		ION BEAM ACTIVATED DEPOSITION	
		HIGH POWER DIODE LASER	
		ICE PREVENTION	

Fig. 1. Proposed ARI reviews.

UNIVERSITIES		GOVERNMENT		INDUSTRY
YALE-2	COLUMBIA-2	ONR-3	NBS-5	BELL LABS-3
PENN-2	BROWN	ONT-7	NIH-2	DUPONT
CAL/BERKLEY	MIT-2	OPNAV-7	NSF	WESTINGHOUSE
JOHN HOPKINS-2	ILLINOIS	USNA	EPA	GENERAL ELECTRIC
NEW YORK U	ST UNIV N.Y.	ARO	DOE-2	OAK RIDGE NAT'L LAB
PENN STATE	BKLYN POLY	AFOSR	NOAA	LINCOLN LABS
MARYLAND-5	RHODE ISLAND	DARPA	NASA	NAT'L CENT ATM RES
SO. CAROLINA	VIRG. POLY			WOODS HOLE O. I.
DREXEL	TEXAS A&M			
GEORGE WASH U	NORTH CAR. ST.			

Fig. 2. Sources of reviewers.

and the Chief of Naval Operations staff. Sources of reviewers for the spring 1985 evaluation are shown in Fig. 2. The panel members are sent background technical information on each proposed program, and then meet in mid-spring to perform their evaluations of the proposed programs. At the panel meeting, a representative of each claimant proposing a new program presents the proposed program to the panel members. These panelists are required to perform a written evaluation of each proposal. A typical agenda is shown on Fig. 3, and copies of the written evaluation form and definitions used in spring 1985 are shown in Figs. 4 and 5. By the end of each review day, all the programs reviewed by the panel which met on that day are ranked in priority order and evaluation forms are completed for each program by each panel member. When all the panels have completed their evaluations, the panel scores for each program, the claimant priorities, and the written comments on each program are analyzed in detail. After ONR requirements are incorporated, a prioritized list of all proposed programs is developed, and this list is presented to the ONR Corporate Board (high-level Navy technical managers who advise the Chief of Naval Research on research-related issues) in late spring for final approval. Approved programs will have funding initiated approximately 15 months after Corporate Board approval at

the start of the fiscal year, i.e., programs approved in June 1985 were funded starting in October 1986.

B. Overview of Evaluation of Existing ARI Programs

Nominally, the programs are funded for a period of five years. Then, all or a part of these programs are transitioned (funding responsibility transferred) to another funding organization, or transitioned to the core, or terminated (or form the basis for a follow-on ARI program). Typically, after two or three years of funding, ONR performs a midterm evaluation of each program. The following section describes the initial evaluation of some of the existing programs using the revised ONR process by the Planning and Assessment Office of ONR which occurred during winter 1984-1985. Since the present method for evaluating proposed programs was initiated in spring 1984, and the present method for evaluating existing ARI programs was initiated in winter 1984-1985, the evaluation processes are somewhat similar. Obviously, the major difference is that performance and team quality play a larger role in determining the quality of existing programs compared to proposed programs. Overall, existing programs were selected for review, were

0830 PRE-BRIEFING - REVIEWERS ONLY - KOSTOFF
 0900 Supersonic Combustion - CRP - Dr. Whitehead
 1000 Solid - Fluid Interfaces - CRP - Dr. Kushner
 1100 Break
 1115 Cavitation - NAVSEA - Dr. Hong
 1215 Ocean Floor Embedment Mechanics - NAVFAC - Mr. Albertsen
 1315 Lunch
 1345 Micromechanics - NAVSEA - Dr. Zanis
 1445 High Strength Ferrous Alloys - NRL - Dr. Vandermeer
 1545 Break
 1600 Executive Session
 1900 Adjourn

Dr. Ronald N. Kostoff - Chairman
 Office of Naval Research
 Arlington, VA

Fig. 3. Mechanics and materials review—April 30, 1985—Room 915:
 Agenda.

Program	Reviewer	Date of Review
IA. RESEARCH MERIT (CIRCLE ONE NUMBER OR -)		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
IB. RESEARCH APPROACH/PLAN/FOCUS/COORDINATION		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
IC. MATCH BETWEEN RESOURCES AND OBJECTIVES		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
ID. BALANCE BETWEEN EXPERIMENT AND THEORY		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
IE. STATE-OF-THE-ART OF EQUIPMENT, THEORIES, AND ALGORITHMS		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
IF. POTENTIAL PERFORMERS/MANAGEMENT TEAM		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
2A. Problem or need(s) to which this research relates.		
2B. PROBABLE IMPACT ON NAVY NEEDS (RESEARCH/TECHNOLOGY/OPERATIONS)		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
2C. POTENTIAL FOR TRANSITION OR UTILITY		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
2D. PHASE OF R&D		
6.1 - 6.2 - 6.3	**BASIC RES.**	APPLIED RES. **EXPLORATORY DEV. ****ADVANCED DEV.****
3. Reviewer's expertise in the research area covered by this program.		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
4. OVERALL PROGRAM EVALUATION		
1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10	***LOW***	****FAIR****
****AVERAGE****		
****GOOD****		
HIGH		
(over) [comments on other side]		

Fig. 4. Science and Mission Impact questionnaire.

The scoresheet contains factors generally related to science issues and Navy relevance issues. It also requires estimates of the reviewer's expertise in the science area, and an overall program score. The scoring bands for all questions except 2A and 2D are shown in 1A: 1-2 (LOW) signifies poor or unacceptable; 2.5-4 (FAIR) is marginally acceptable; 4.5-6.5 (AVERAGE) is passable or acceptable; 7-8.5 (GOOD) represents better than average; 9-10 (HIGH) is outstanding, top quality. Question 2A does not require a numerical response, and question 2D has its scoring range defined.

DEFINITIONS

- 1A. RESEARCH MERIT-Importance to science of question or problem addressed by the program. Consider definition of objectives, advancement of state-of-art, uniqueness of contribution, and degree of fundamentality.
- 1B. RESEARCH APPROACH/PLAN/FOCUS/COORDINATION-Process employed to solve the research problem, including the quality and focus of the research plan, degree of innovation, understanding of field, and coordination with, or cognizance of, other related programs to minimize duplication or gaps.
- 1C. MATCH BETWEEN RESOURCES AND OBJECTIVES-Relationship between scientific objectives proposed and total resources requested.
- 1D. BALANCE BETWEEN EXPERIMENT AND THEORY-Adequacy of experiment/theory balance to accomplish proposed scientific objectives.
- 1E. STATE-OF-THE-ART OF EQUIPMENT, THEORIES, AND ALGORITHMS-Degree of advancement of tools proposed to solve research problem.
- 1F. POTENTIAL PERFORMERS/MANAGEMENT TEAM-Quality of principal researchers on program (if known) and, if applicable, management team.
- 2A. Description of need, either research and/or technology and/or operational, to which this research relates.
- 2B. PROBABLE IMPACT ON NAVY NEEDS (RESEARCH/TECHNOLOGY/OPERATIONS)-Potential impact of this program on Navy research/technology/operational needs multiplied by probability of success.
- 2C. POTENTIAL FOR TRANSITION OR UTILITY-Probability that results of this program will be transitioned to or utilized by Navy technical community.
- 2D. PHASE OF R&D-Level of development of the program. Scale ranges from basic research (6.1) through exploratory development (6.2) to advanced development (6.3).
- 4. OVERALL PROGRAM EVALUATION-Single number description of overall program quality based on all relevant criteria. Provide narrative of pros and cons and any recommendations under COMMENTS.

Fig. 5. Scoring criteria.

reviewed individually, and the results were sent to the appropriate claimant. Where necessary, redirections were specified.

The review process for existing programs starts with the generation of rating factors and criteria (such as team quality, research merit, etc.) with which the reviewers will evaluate the program. Once the rating factors have been defined, and specific programs have been targeted for review, then the appropriate panel members are selected. The program presenters and reviewers are provided with technical information about the program. The reviewers assemble, hear presentations about the program, and perform written evaluations followed by a closed-session discussion about the program. The chairman of the review group writes a report on the program evaluation, and this report and recommendations are sent to the program managers.

III. DETAILED DESCRIPTION OF REVIEW PROCESSES FOR EXISTING AND PROPOSED ARI PROGRAMS

The major components of the present review process are arbitrarily categorized in chronological form as Development of Rating Criteria, Selection of Reviewers, Priming Presenters, Priming Reviewers, Conduct of Presentation, Conduct of Evaluation, and Final Report and Feedback. These major components will now be discussed in detail. For those individuals who are interested in implementing the present review process in their organization, it is highly recommended that they read this section carefully. This section contains many important details that have been developed with years of experimentation and which contribute heavily to the successful operation of the evaluation process.

A. Development of Rating Criteria and Factors

The review criteria and rating factors chosen impact the conduct of the review as well as the selection of reviewers, and choice of these criteria and rating factors should be based on evidence of their importance to the reviewers' determination of quality of research. Providing definitions of the evaluation criteria is a step toward normalizing and standardizing the review process. In addition, another step toward common interpretation of results is definition of scoring bands. To minimize reviewer confusion and unnecessary delay during the short written evaluation period, the scoring bands for all rating factors should be similar (identical if possible).

In the evaluation of existing programs which occurred in Jan./Feb. 1985, ONR utilized criteria and rating factors which had been tested and validated for the past four years both in ONR and DOE (See Fig. 6 for a list of rating factors). The criteria used address the major issues of science and Navy relevance. The four rating factors in Questions 1A-1D are those which hundreds of reviewers have considered most important when evaluating the science quality of a program, and the rating factor in Question 2B was deemed important in the evaluation of mission relevance. The importance of these specific rating factors in the reviewers' determination of the Overall Program Quality rating factor was identified by reading hundreds of reviewers' comments in previous ONR and DOE evaluations and grouping these responses into key rating factors, and by the regression analysis summarized in the Appendix and described in detail in [1]. The Phase of R&D rating factor in Question 2C had a somewhat unique history. In the evaluation of proposed programs

Program _____

Reviewer _____ Date of Review _____

1A. QUALITY OF PROJECT TEAM (CIRCLE ONE NUMBER OR -)
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 An outstanding team would rate 9-10, a strong balanced team of experienced investigators would rate 6-8, a team which would profit from some strengthening would rate 3-5, and a team with serious shortcomings would rate 0-2.
 COMMENTS: _____

1B. RESEARCH MERIT
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 A problem of central importance to advancing the research area would rate 9-10, sound programmatic research providing information of general usefulness and interest would rate 6-8, a routine research problem would rate 3-5, and a problem of doubtful or peripheral interest would rate 0-2.
 COMMENTS: _____

1C. SCIENTIFIC APPROACH, TECHNIQUES, RESOURCES
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 A skillful and innovative approach would rate 9-10, a creditable and logical use of state-of-the-art methods would rate 6-8, an out-of-date approach would rate 3-5, and an inappropriate or illogical approach would rate 0-2.
 COMMENTS: _____

1D. PRODUCTIVITY
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 Extensive, high quality output would rate 9-10, interesting results at a reasonable pace would rate 6-8, marginal output would rate 3-5, and little evidence of progress would rate 0-2.
 COMMENTS: _____

2A. Indicate the problem or need(s) to which this research relates.

2B. Probable impact (potential impact x probability of success) of this project on Navy needs (research and/or technology and/or operations).
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 *****LOW***** *****MEDIUM***** *****HIGH*****

2C. Phase of R&D of this project. (CIRCLE ONE NUMBER or \$)
 6.1 \$ \$ \$ \$ 6.2 \$ \$ \$ 6.3
 BAS.RES. ***APP.RES.*** ***EXP.DEV.*****ADV.DEV.*****

3. Estimate your expertise in the research area covered by this project.
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 *****LOW***** *****MEDIUM***** *****HIGH*****

4. OVERALL PROJECT EVALUATION
 0 - 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10
 An outstanding project would rate 9-10, a strong project deserving of continuing priority would rate 6-8, a project with some deficiencies or concerns would rate 3-5, and a project with serious shortcomings which warrants close re-evaluation would rate 0-2.
 COMMENTS: (PROS & CONS, AND RECOMMENDATIONS) _____

Fig. 6. Science and Mission Impact questionnaire.

which occurred in spring 1984, Appropriateness (for ONR funding) was introduced as a rating factor based on the comments of reviewers in previous ONR and DOE evaluations. However, as a result of the spring 1984 review, Appropriateness was deleted, and Phase of R&D (basic research, applied research, exploratory development, etc.) was added as a factor. It was felt, based on the initial ONR application of the process, that "Appropriateness" had a number of components which most of the external reviewers did not have the organizational understanding to evaluate. However, the one component of "Appropriateness" which would address the tie between the program and ONR's science mission, and which it was felt the reviewers were eminently qualified to evaluate was "Phase of R&D," and for this reason the rating factor was added. ONR provided the reviewers with the criteria and rating factor definitions as shown in Fig. 7 and defined scoring bands for each rating factor as shown in Fig. 6.

The criteria and rating factors for proposed programs are somewhat similar to those used for existing programs. Productivity was not used since the programs hadn't started, and scoring the Potential Performers criterion (which was substituted for the Quality of Project Team criterion) was optional since in many broad-based programs the performers would be known only after detailed

proposals were received and approved. Scientific Approach, which had been used as a separate rating factor in the Jan./Feb. 1985 evaluation of existing programs, was subdivided into Research Approach/Plan/Focus/Coordination, Match Between Resources and Objectives, Balance Between Experiment and Theory, and State of the Art of Equipment, Theories, and Algorithms rating factors for the evaluation of proposed programs. A careful reading of reviewers' comments in previous evaluations showed that reviewers were making this subdivision in order to be able to evaluate the Scientific Approach rating factor fully. Potential for Transition or Utility was added as a rating factor. Future reviews of existing programs will use the formats and criteria described in this paragraph for proposed programs, with the necessary additions of Productivity and Performer Quality rating factors. ONR provided the reviewers with the rating factor definitions as shown in Fig. 5, and defined identical scoring bands for each major factor as shown in Fig. 4.

B. Selection of Reviewers

As the number of reviewers increases, more depth of expertise as well as diversity of opinion due to a better statistical sampling of the relevant technical community can be obtained, and it becomes easier

The scoresheet (Attachment 2) contains criteria generally related to science issues and Navy impact issues. It also requires estimates of the reviewer's expertise in the subject area and an overall project score.

For item 1A, Quality of Project Team, the reviewer should consider the quantity and quality of publications by the principal scientist, honors and awards, review articles, experience relevant to the project area, and other less tangible factors which combine to provide a reputable and high quality team.

For item 1B, Research Merit, the reviewer should assess the importance of the scientific question or problem addressed by the project, and consider such issues as definition of objectives, advancement of state-of-the-art, uniqueness of contribution, and degree of fundamentality. This judgement should be based primarily on the reviewer's personal knowledge of the scientific and engineering fields, as well as his knowledge of the results of other review panels and workshops.

For item 1C, Scientific Approach, Techniques, Resources, the reviewer should assess the total process employed to solve the research problem, including the quality of the planning, match between resources and objectives, balance between experiment and theory, degree of innovation, understanding of field, and state-of-the-art of techniques and equipment.

For item 1D, Productivity, the reviewer should assess the volume and quality of work produced and relate this output to the resources available and costs incurred. Projects which have been initiated recently should not be rated on productivity.

Item 2A is the only question on the questionnaire which does not require a numerical answer. It asks the reviewer to identify the need to which the research is related, and thus requires the reviewer to articulate the relationship of the scientific issues associated with the project's subject area to the agenda of the project.

For item 2B, Probable Impact, the reviewer should estimate the probable impact of the project on the needs of the Navy. To evaluate probable impact, the reviewer should estimate potential impact of this project on Navy needs (if successful), and then multiply this number by his estimate of probability of success. For purposes of this review, the broadest concept of Navy needs should be utilized to include Navy needs in research, technology, and operations.

For item 2C, Phase of R&D, the reviewer should estimate the level of development of the project. The most fundamental research is at the left-hand side of the scale (6.1) and advanced development is at the right-hand side (6.3). The dollar signs (\$) represent divisions along the scale. Applied research ranges from about 6.15 to 6.2, and exploratory development occupies most of the range from 6.2 to 6.3.

For item 4, Overall Project Evaluation, the reviewer should select one number which he feels characterizes the quality of the overall project. The reviewer should take into account not only the items evaluated above but any other relevant criteria as well. Under COMMENTS in item 4, the reviewer should describe in narrative form the pros and cons of this project, and any recommendations which he may have for changing this project. Whenever possible, detailed comments or recommendations are preferred.

Fig. 7. Scoring criteria.

to separate between common judgements among reviewers and personal biases. On the other hand, as the number of reviewers increases, the cost of the review becomes higher, the complexity of the review arrangements becomes greater in a panel process, and the actual conduct of the review becomes more unwieldy.

Credible reviews require that there be sufficient expertise on each panel to address each rating factor for each program under consideration. Different types of reviewers in the same review bring different perspectives and provide a more synergistic view of the program in its overall Navy context. Organization and personal reviewer conflicts and biases can have major impacts on the outcome of the reviews. Reviewer conflicts should be a prime motivator for disqualifying reviewers.

In the single program per panel format (only one program was reviewed by each panel) that was used in the initial ONR application of the revised process for existing programs in Jan./Feb. 1985, nominally six reviewers per program were used, which appeared to be a good trade-off among diversity, complexity, and manageability. Reviewers were selected such that half were experts in the science area of the program being evaluated, and the other half were experts in the Navy relevance aspects. Specifically, the science experts tended to be senior level researchers and/or technical managers in the field of the program. The relevance experts typically consisted of

representatives of the Navy Exploratory Development Program and an OPNAV (Chief of Naval Operations) representative to address the Navy requirements issue. Thus, each rating factor was covered by at least one (and usually two or three) credible expert. These science experts and relevance experts were combined into one panel to provide cross-fertilization among disciplines and to elucidate the many dimensions of the program in the Navy role. Claimants were excluded as reviewers.

The proposed programs were reviewed in a multiprogram per panel format (each panel reviewed more than one proposal). The number of reviewers on each panel depended on the number of programs being evaluated. The number of programs reviewed by any one panel ranged between four and six during the one-day panel meeting, and the number of reviewers per panel ranged between ten and twelve. For each program being evaluated, there was at least one panel member who was an expert in the program's science area. In addition, there tended to be one or two senior scientists of sufficient breadth to cover a number of the similar science programs. The remainder of the panel typically consisted of Navy relevance experts similar to those used on the evaluation panels for existing programs.

If a member of the Navy Research Reserve could be identified who was both a distinguished scientist in a review panel's discipline and a senior Navy reserve officer, this individual was added to the panel to

provide an additional type of capability in science/relevance expertise. Thus, in the multiprogram per panel approach, while there are by necessity fewer detailed experts per program than in the single program review per panel approach, there is a broader range of expertise on the panel which allows a better comparison of the proposed programs by one group of reviewers. In addition to the cross-fertilization due to having representatives from different phases of development as described above, there were additional perspectives provided by the mix of technical specialties within the general science area of the panel. To minimize reviewer conflict, a list of candidate reviewers was presented to the claimants before the review, and if a case were made by any claimant that potential bias could be present for a particular reviewer, that reviewer was removed from the list.

C. Priming Presenters

Providing claimant representatives who will present the proposals with evaluation forms and information about the review process beforehand allows them to maximize the relevant information to be provided to the reviewers and to target their presentations to the evaluation factors. To familiarize the reviewers with the program beforehand, and not to have them enter the review process "cold," it is valuable to have the presenters provide them with as much detail as reasonably practical beforehand. A minimum requirement from presenters of existing programs should be a list of the principal researchers on the program along with a brief description of their expertise areas, and a list of outputs from the program to date (papers, patents, equipment, etc.).

In the review of existing programs, presenters were given both time limits for presentation and a list of scoring factors. A program narrative was required from the presenters.

Presenters of proposed programs were given both time limits for presentation and a list of scoring factors, as well as the more detailed evaluation information contained in Figs. 3–5. A brief program narrative was required from the presenters.

D. Priming Reviewers

As information that is provided to the reviewers beforehand about the conduct of the process increases, time expended during the review to explain the process decreases. To eliminate any potential misunderstandings on the part of the reviewers, a description of the process and their role in it should be sent beforehand and a brief verbal summary should be provided on the day of the review. In order for the review to proceed smoothly, it is very important for the types of programs which ONR funds to be differentiated clearly during the verbal summary. When more technical information is provided to the reviewers beforehand, they will be better prepared for incisive questioning when they enter the review.

The reviewers of existing programs were sent a description of the overall process, evaluation sheets, and definitions of scoring criteria. Copies of the narrative were sent before the review and copies of viewgraphs were provided at the review. In the future, narratives, viewgraphs, researchers' descriptions, and specific program outputs will be a minimum requirement of material due before the review.

In the most recent reviews of proposed programs, reviewers received copies of the viewgraphs which the presenters would show at the panel review beforehand, in order for the reviewers to start formulating specific questions well before the review.

The reviewers were sent a description of the overall process, evaluation sheets, and definitions of scoring criteria beforehand, and were provided a brief verbal summary of the process and its context within the ONR program the morning of the review day. Copies of the program narrative were sent beforehand, and copies of the viewgraphs were provided beforehand and at the review.

E. Conduct of Presentation

To allow the reviewers to make credible judgements, sufficient time must be allowed for the presentation and interaction with the audience. Too little time restricts the data base on which the reviewers judge the program, and too much time becomes cost-ineffective, and limits the number of programs which can be reviewed in a given day.

For the evaluation of existing programs, only one program per day was reviewed and consequently the total time was allowed to "float." Each presentation was allowed to run its natural course. Thirty-five minutes were allowed nominally, but because of the flexible schedule, some presenters went well beyond this limit. Twenty-five minutes were allowed nominally for questions and answers, but again this limit was exceeded substantially. The average total presentation time per program was about 1.5–2 h. While this total presentation time is about twice the time experienced in the proposed program presentations, it should be remembered that the existing program reviews include research results. In some cases, discussion of research results was extensive.

For the evaluation of proposed programs, 25 min for presentation and 20 min for questions and answers were allowed nominally per program, and because of the tight multiprogram schedules, most of the presenters were kept within these time guidelines. In more recent reviews, when ONR received viewgraphs beforehand, presentations which appeared too long were discussed with the presenters and overruns were avoided.

F. Conduct of Evaluation

The primary data of the review process is contained in the written comments of the reviewers, and sufficient "quality time" should be allocated to allow intelligible and integrated comments to be provided. There is value in generating the written reviews in the period immediately after the presentation. This is because the subject matter is still fresh in the minds of the reviewers, external influences have not had an opportunity to generate any reviewer biases, the reviewers can concentrate on the evaluation and not be distracted by the speakers and audience, and the comments can form the basis for later discussion. Postwritten review discussion is valuable, in that it allows the opportunity to eliminate any misunderstanding, allows for useful interaction among reviewers in a more "open" atmosphere than the presentation, and allows the possibility of arriving at some kind of consensus view of the program and its components.

A nominal written evaluation time of 15 min was provided to the reviewers of existing programs, but the time was allowed to "float." The average time taken was about 20 min. From the ONR experiment, and the experience of many past reviews, a 15-min period after the presentation is required for a reasonable written evaluation of a specific program, although 10 min would probably be acceptable in a time-constrained environment. Thirty minutes for the discussion period was allowed nominally, but again the time was allowed to "float." The median discussion time taken was about 45 min, but discussion time for all the existing programs evaluated ranged from about 30 min to 2 h. Based on the ONR experiments, and previous reviews, 30 min of discussion appears to be a minimum for a time-constrained focused program review.

For proposed programs, 15 min for initial written evaluations after each presentation were allowed. Thirty minutes for discussion of each program were allowed at the end of the review day, and because of the time constraints the reviewers were held closely to this schedule. At the start of the discussion period, reviewers were provided with their scoresheets and were given the opportunity to modify their scores during this discussion time, as well as to modify their written comments. A consensus score on each rating factor was reached for

each program. Then, the consensus Overall Program Quality scores were compared for the programs which the panel reviewed, and if the panel was not comfortable with the relative rankings, it was allowed to modify the consensus Overall Program Quality scores to produce a more acceptable prioritization.

G. Final Report and Feedback

A written report of the review which documents the process and the reviewing organization's evaluation and proposed actions provides a solid basis for discussion and is valuable for future reference. The "bottom line" of the program review process of existing programs is the feedback and redirection (if required) provided to the claimant in order to improve the program's quality.

ONR produced a written report for each existing program reviewed which incorporated the reviewers' written comments, conclusions reached during the Executive Discussion, and the chairman's recommendations based on reviewers' comments and discussion. Reviewers' written comments (with reviewer identification deleted) were included verbatim as an Appendix. ONR provided verbal feedback of each review's results to the program manager for each program reviewed. When appropriate, based on reviewer recommendations, the technical director of ONR sent each written report to the relevant claimant and requested modifications to the reviewed program.

A one-page summary document which included reviewers' key written comments and major comments from the Executive Discussion was written for each proposed program. This summary document and reviewers' verbatim comments (with reviewer identification deleted) were sent to each proposer. The reviewers' scores were tabulated and analyzed, and along with the summary document, were presented to the ONR Corporate Board with a prioritization of the proposed programs and funding recommendations.

IV. APPLICATIONS OF THE REVISED ONR EVALUATION PROCESS

A. Existing ARI Programs

In Jan./Feb. 1985, five panels were formed to evaluate five existing ONR research programs. Smaller panels were used relative to those which evaluated proposed programs (six reviewers as opposed to twelve reviewers in the multiprogram reviews), but the same combination of researchers, transitioners, technologists, and requirements experts was retained. The time flexibility using the one program per panel review format allowed experiments to be run in which review parameters such as length of presentation, length of written evaluation, and length of discussion session were varied. The knowledge gained from these experiments was extremely helpful in designing subsequent reviews. Since only one program was evaluated by each panel, it was decided to restrict selection of reviewers to the local commuting area. Total cost to the evaluating organization for all five panels, excluding staff time, was under \$200.

B. Proposed ARI Programs

The revised evaluation process has been applied to proposed research programs in April 1984, 1985, 1986, and 1987. In contrast to previous ONR practice, a large number of the panel members (mainly the science experts) were selected from outside of ONR and the Navy. It was hoped that having reviewers external to, and independent from, ONR and the Navy would lead to a greater degree of objectivity in the review process. The panel rankings have proved to be the major factor in determining the final programs recommended for ONR funding.

For each of the four years, the number of panels ranged from five to seven, and the number of research programs evaluated ranged from about twenty-five to thirty. The first year of funding for

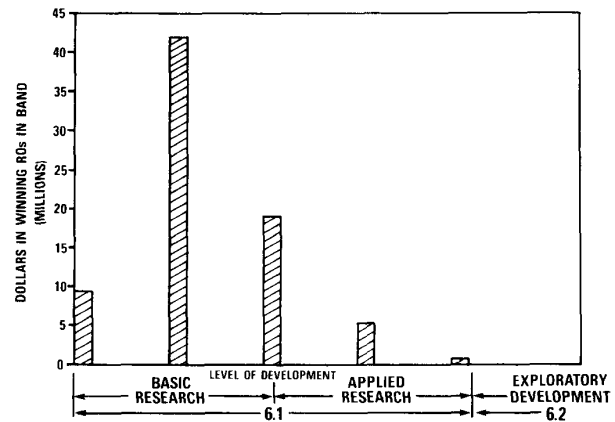


Fig. 8. Dollars in winning RO's versus R&D phase.

programs which were successful in the competition for each of the four years ranged from \$20 million to \$30 million, and the ratio of winning programs to proposed programs ranged from 0.6 to 0.8, depending on the funds in the competition pool. The total cost of the evaluation to ONR, excluding ONR staff time, ranged each year from \$7000 to \$20 000. These expenditures were predominantly for reviewers' travel and lodging. The highest review cost resulted from a decision to broaden geographical representation of the outside reviewers, with the attendant higher travel costs.

V. ANALYSIS OF REVIEWERS' SCORES

In fall 1987, extensive analyses were performed on the panel reviewers' written scores for programs proposed in spring 1985, 1986, and 1987. One of the more interesting results follows.

Fig. 8 is a plot of dollar frequency distribution as a function of the phase of R&D criterion discussed earlier for those programs which won the competition and will receive funding. The funding distribution function for phase of R&D peaks in the middle of the basic research range, as would be expected for a mission-oriented program. Eighty-eight percent of these programs (by funding level) are judged to be in the basic research phase, and the remainder are judged to be in the applied research phase.

VI. CONCLUSION AND DISCUSSION

ONR has developed a process which is sufficiently flexible to evaluate existing and proposed research programs. The method is low cost for the evaluation organization; the costs for the organization being evaluated have not been estimated and could span a wide range, depending on the preparation effort by the reviewees. The method insures that the rating factors which determine high quality are used during the evaluation process through the use of multiple regression and other supporting analyses, and that the appropriate credible experts are used to evaluate each scoring factor for each program/project reviewed. The method incorporates substantial interaction among the reviewers to insure that minimal misconceptions remain when a consensus panel score is obtained.

The process has been used to evaluate proposed and existing accelerated research programs which constitute about 40 percent of the total ONR budget. This process is starting to be used, in part at least, by a number of the Navy claimants to evaluate internally their funded programs. In addition, a number of panel reviewers from outside ONR who have participated in the ONR evaluations of proposed and existing research programs have stated that they plan to incorporate this process, in whole or in part, in their own organizations' research evaluation cycle.

For organizations wishing to adapt this process to internal program evaluations, development and validation of review criteria will be a primary concern. Review criteria should be related to program goals. The general ONR approach for development and validation of review criteria should be applicable to any other organization.

In this approach, criteria were identified to insure that all key issues relating to program quality in the context of Navy goals for research were addressed. The criteria were tested in a program evaluation cycle and, based on reviewers' supplementary written comments about the programs, criteria were modified to insure major reviewer concerns would be addressed by the criteria in future evaluations. In parallel, a regression approach similar to that used in [1] was employed to determine whether the rating factors used were the major determinants of the reviewers' Overall Program Evaluation scores.

APPENDIX

DOE EVALUATION PROCESS FOR EXISTING RESEARCH PROJECTS DEVELOPED IN 1981

In 1981, the DOE decided to evaluate existing research projects funded by its Office of Basic Energy Sciences (BES). The DOE used a rather innovative panel-based peer review approach [1], [2]. Out of approximately 1200 active projects supported by BES, a randomly selected sample of 129 projects was reviewed by panels of scientific peers. The projects were grouped by areas of similar science, and the reviews were conducted on 40 separate days by 40 separate expert panels, with an average of four members and three projects per panel. The reviewers were, for the most part, bench level scientists independent of the DOE.

On a given review day, the projects of similar science would be evaluated by the one panel responsible for the technical area. The principal investigator for a given project would make a presentation, and the reviewers were then given a period of time to evaluate the project by scoring a number of factors and generating written comments to support their scores. After all the projects had been presented and scored, the reviewers assembled in an Executive Session, discussed each project in detail, and arrived at a consensus panel score for each project.

The reviewers were asked to evaluate seven rating factors independently for each project: Team Quality, Scientific Merit, Scientific Approach, Productivity, Importance to Mission, Energy Impact, and Overall Project Quality. These rating factors were selected based mainly on the author's past experience with project reviews; there was essentially no data to support the choice of these rating factors as the seven "best," or "most complete," or any other similar descriptor. After the scoring by the panels was completed, a six-factor multiple-regression model was used to relate the Overall Program Quality rating factor to the other six rating factors. The six factor model produced a correlation coefficient of 0.89, which meant that the six factors selected constituted the bulk of the considerations which the reviewers used to score the Overall Project Quality rating factor. In fact, the best three-factor model derived to predict the Overall Project Quality rating factor score consisting of Team Quality, Scientific Approach, and Importance to Mission produced correlation coefficients within 3 percent of those obtained with the complete six-factor model [1].

The cost of the DOE evaluation process was not negligible. According to Logsdon, [2], who made an independent assessment of the DOE (as well as other federal agency research) evaluation process, "the average cost per project evaluated is approximately \$7000." This number represents cost to the evaluating organization only, excluding staff time. Since the 129 projects evaluated had a mean funding level of about \$150 000 per year, this meant that almost \$1 million, excluding staff time, was spent by the evaluating organization alone to evaluate directly the research quality of about \$20 million worth of existing research projects and to infer the research quality of the total BES program. The costs to BES and the principal investigators (the reviewees) were not estimated.

REFERENCES

- [1] Office of Program Analysis, Office of Energy Research, "An assessment of the basic energy sciences program," U.S. Department of Energy, Publ. DOE/ER-0123, Mar. 1982.
- [2] J. M. Logsdon, "An overview of Federal research evaluation activities," Program of Policy Studies in Science and Technology, George Washington Univ., Washington, DC, Rep., Apr. 1985.